PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/00 | A2

(11) International Publication Number: WO 99/24463

(43) International Publication Date: 20 May 1999 (20.05.99)

(21) International Application Number: PCT/US98/23578

(22) International Filing Date: 4 November 1998 (04.11.98)

(30) Priority Data:
08/966,316 7 November 1997 (07.11.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/966,316 (CIP)
Filed on 7 November 1997 (07.11.97)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). MURRY, Lynn, E. [US/US]; 1124 Los Trancos Road, Portola Valley, CA 94028 (US). MATHUR, Preete [US/US]; 43733 Greenhills, Fremont, CA 94539 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SIGNAL PEPTIDE-CONTAINING PROTEINS

```
      9          18          27          36          45          54
NGC GAC GTA CAA CAG ATT GGA GCC ATG GCT TTG GAA CAG AAC CAG TCA ACA GAT
                                M   A   L   E   Q   N   Q   S   T   D

         63          72          81          90          99         108
TAT TAT TAT GAG GAA AAT GAA ATG AAT GGC ACT TAT GAC TAC AGT CAA TAT GAA
Y   Y   Y   E   E   N   E   M   N   G   T   Y   D   Y   S   Q   Y   E

        117         126         135         144         153         162
CTG ATC TGT ATC AAA GAA GAT GTC AGA GAA TTT GCA AAA GTT TTC CTC CCT GTA
L   I   C   I   K   E   D   V   R   E   F   A   K   V   F   L   P   V

        171         180         189         198         207         216
TTC CTC ACA ATA GTT TTC GTC ATT GGA CTT GCA GGC AAT TCC ATG GTA GTG GCA
F   L   T   I   V   F   V   I   G   L   A   G   N   S   M   V   V   A

        225         234         243         252         261         270
ATT TAT GCC TAT TAC AAG AAA CAG AGA ACC AAA ACA GAT GTG TAC ATC CTG AAT
I   Y   A   Y   Y   K   K   Q   R   T   K   T   D   V   Y   I   L   N

        279         288         297         306         315         324
TTG GCT GTA GCA GAT TTA CTC CTT CTA TTC ACT CTG CCT TTT TGG GCT GTT AAT
L   A   V   A   D   L   L   L   L   F   T   L   P   F   W   A   V   N
```

(57) Abstract

The invention provides signal peptide–containing proteins collectively designated SP, and polynucleotides which identify and encode these molecules. The invention also provides expression vectors, host cells, agonists, antibodies and antagonists. The invention further provides methods for diagnosing, treating, and preventing disorders associated with expression of signal peptide–containing proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SIGNAL PEPTIDE-CONTAINING PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of new signal peptide-containing proteins which are important in disease and to the use of these sequences in the diagnosis, treatment, and prevention of diseases associated with cell proliferation and cell signaling.

BACKGROUND OF THE INVENTION

Protein transport is a quintessential process for both prokaryotic and eukaryotic cells. Transport of an individual protein usually occurs via an amino-terminal signal sequence which directs, or targets, the protein from its ribosomal assembly site to a particular cellular or extracellular location. Transport may involve any combination of several of the following steps: contact with a chaperone, unfolding, interaction with a receptor and/or a pore complex, addition of energy, and refolding. Moreover, an extracellular protein may be produced as an inactive precursor. Once the precursor has been exported, removal of the signal sequence by a signal peptidase activates the protein.

Although amino-terminal signal sequences vary substantially, many patterns and overall properties are shared. Recently, hidden Markov models (HMMs), statistical alternatives to FASTA and Smith Waterman algorithms, have been used to find shared patterns, specifically consensus sequences (Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197). Although they were initially developed to examine speech recognition patterns, HMMs have been used in biology to analyze protein and DNA sequences and to model protein structure (Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Collin, M. et al. (1993) Protein Sci. 2:305--314). HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides and for opening and extending an insertion or deletion. The algorithms are quite flexible in that they incorporate information from newly identified sequences to build even more successful patterns. To find signal sequences, multiple unaligned sequences are compared to identify those which encode a peptide of 20 to 50 amino acids with an N-terminal methionine.

Some examples of the protein families which are known to have signal sequences are receptors (nuclear, 4 transmembrane, G protein coupled, and tyrosine kinase), cytokines (chemokines), hormones (growth and differentiation factors), neuropeptides and vasomediators, protein kinases, phosphatases, phospholipases, phosphodiesterases, nucleotide cyclases, matrix molecules (adhesion, cadherin, extracellular matrix molecules, integrin, and selectin), G proteins, ion channels (calcium, chloride, potassium, and sodium), proteases, transporter/pumps (amino acid, protein, sugar, metal and vitamin; calcium, phosphate, potassium, and sodium) and regulatory proteins. Descriptions of some of these proteins (receptors, kinases, and matrix proteins) and diseases associated with their dysfunction follow.

G-protein coupled receptors (GPCR) are a large group of receptors which transduce extracellular signals. GPCRs include receptors for biogenic amines such as dopamine, epinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin; for lipid mediators of inflammation such as prostaglandins, platelet activating factor, and leukotrienes; for peptide hormones such as calcitonin, C5a anaphylatoxin, follicle stimulating hormone, gonadotropin releasing hormone, neurokinin, oxytocin, and thrombin; and for sensory signal mediators such as retinal photopigments and olfactory stimulatory molecules. The structure of these highly-conserved receptors consists of seven hydrophobic transmembrane regions, an extracellular N-terminus and a cytoplasmic C-terminus. The N-terminus interacts with ligands and the C-terminus interacts with intracellular G proteins to activate second messengers such as cyclic AMP (cAMP), phospholipase C, inositol triphosphate, or ion channel proteins. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved, acidic-Arg-aromatic triplet present in the second cytoplasmic loop may interact with the G proteins. The consensus pattern, [GSTALIVMYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM] is characteristic of most proteins belonging to this group (Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego, CA; Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10).

The kinases comprise the largest known group of proteins, a superfamily of enzymes with widely varied functions and specificities. Kinases regulate many different cell

proliferation, differentiation, and signaling processes by adding phosphate groups to proteins. Receptor mediated extracellular events trigger the transfer of these high energy phosphate groups and activate intracellular signaling cascades. Activation is roughly analogous to the turning on a molecular switch, and in cases where signalling is uncontrolled, may be associated with or produce inflammation and cancer.

Kinases are usually named after their substrate, their regulatory molecule, or after some aspect of a mutant phenotype. Almost all kinases contain a similar 250-300 amino acid catalytic domain. The N-terminal domain, which contains subdomains I-IV, generally folds into a two-lobed structure which binds and orients the ATP (or GTP) donor molecule. The larger C terminal lobe, which contains subdomains VIA-XI, binds the protein substrate and carries out the transfer of the gamma phosphate from ATP to the hydroxyl group of a serine, threonine, or tyrosine residue. Subdomain V spans the two lobes.

The kinases may be categorized into families by the different amino acid sequences (between 5 and 100 residues) located on either side of, or inserted into loops of, the kinase domain. These amino acid sequences allow the regulation of each kinase as it recognizes and interacts with its target protein. The primary structure of the kinase domain is conserved and contains specific residues and identifiable motifs or patterns of amino acids. The serine threonine kinases represent one family which preferentially phosphorylates serine or threonine residues. Many serine threonine kinases, including those from human, rabbit, rat, mouse, and chicken cells and tissues, have been described (Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Books, Vol I:7-20 Academic Press, San Diego, CA).

The matrix proteins (MPs) provide structural support, cell and tissue identity, and autocrine, paracrine and juxtacrine properties for most eukaryotic cells (McGowan, S.E. (1992) FASEB J. 6:2895-2904). MPs include adhesion molecules, integrins and selectins, cadherins, lectins, lipocalins, and extracellular matrix proteins (ECMs). MPs possess many different domains which interact with soluble, extracellular molecules. These domains include collagen-like domains, EGF-like domains, immunoglobulin-like domains, fibronectin-like domains, type A domain of von Willebrand factor (vWFA)-like modules, ankyrin repeat modules, RDG or RDG-like sequences, carbohydrate-binding domains, and calcium ion-binding domains.

For example, multidomain or mosaic proteins play an important role in the diverse functions of the ECMs (Engel, J. et al. (1994) Development S35-42). ECM proteins

(ECMPs) are frequently characterized by the presence of one or more domains which may contain a number of potential intracellular disulphide bridge motifs. For example, domains which match the epidermal growth factor tandem repeat consensus are present within several known extracellular proteins that promote cell growth, development, and cell signaling. Other domains share internal homology and a regular distribution of single cysteines and cysteine doublets. In the serum albumin family, cysteine arrangement generates the characteristic 'double-loop' structure (Soltysik-Espanola, M. et al. (1994) Dev. Biol. 165:73-85) important for ligand-binding (Kragh-Hansen, U. (1990) Danish Med. Bull. 37:57-84). Other ECMPs are members of the vWFA-like module superfamily, a diverse group of proteins with a module sharing high sequence similarity. The vWFA-like module is found not only in plasma proteins but also in plasma membrane and ECMPs (Colombatti, A. and Bonaldo, P. (1991) Blood 77:2305-2315). Crystal structure analysis of an integrin vWFA-like module shows a classic "Rossmann" fold and suggests a metal ion-dependent adhesion site for binding protein ligands (Lee, J.-O. et al. (1995) Cell 80:631-638).

The diversity, distribution and biochemistry of MPs is indicative of their many, overlapping roles in cell proliferation and cell signaling. MPs function in the formation, growth, remodeling, and maintenance of bone, and in the mediation and regulation of inflammation. Biochemical changes that result from congenital, epigenetic, or infectious diseases affect the expression and balance of MPs. This balance, in turn, affects the activation, proliferation, differentiation, and migration of leukocytes and determines whether the immune response is appropriate or self-destructive (Roman, J. (1996) Immunol. Res. 15:163-178).

Adenylyl cyclases (AC) are a group of second messenger molecules which actively participate in cell signaling processes. There are at least eight types of mammalian ACs which show regions of conserved sequence and are responsive to different stimuli. For example, the neural-specific type I AC is a Ca^{++}-stimulated enzyme whereas the human type VII is unresponsive to CA^{++} and responds to prostaglandin E1 and isoproterenol. Characterization of these ACs, their tissue distribution, and the activators and inhibitors of the different types of ACs is the subject of various investigations (Nielsen, M.D. et al. (1996) J. Biol. Chem. 271:33308-16; Hellevuo, K. et al. (1995) J. Biol. Chem. 270:11581-9). AC interactions with kinases and G proteins in the intracellular signaling pathways of all tissues make them interesting candidate molecules for pharmaceutical research.

ATP diphosphohydrolase (ATPDase) is an enzyme expressed and secreted by quiescent endothelial cells and involved in vasomediation. The physiological role of ATPDase is to convert ATP and ADP to AMP. When this conversion occurs in the blood vessels during inflammatory response, it prevents extracellular ATP from causing vascular injury by inhibiting platelet activation and modulating vascular thrombosis (Robson, S.C. et al. (1997) J. Exp. Med.185:153-63).

The discovery of new signal peptide-containing proteins and the polynucleotides encoding these molecules satisfies a need in the art by providing new compositions useful in the diagnosis, treatment, and prevention of diseases associated with cell proliferation and cell signaling, particularly cancer, immune response and neuronal disorders.

SUMMARY OF THE INVENTION

The invention features a substantially purified signal peptide-containing protein (SP) having an amino acid sequence selected from the group encoded by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

The invention further provides isolated and substantially purified polynucleotide sequences encoding SP. In a particular aspect, the polynucleotide has a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

In addition, the invention provides a polynucleotide sequence, or fragment thereof, which hybridizes to any of the polynucleotide sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17. In another aspect, the invention provides a composition comprising isolated and purified polynucleotide sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17, or a fragment thereof.

One aspect of the invention features an isolated and substantially purified polynucleotide which encodes SP-16. In a particular aspect, the polynucleotide is the nucleic acid sequence of SEQ ID NO:17. In another aspect, the polynucleotide is a fragment or an oligonucleotide comprising the nucleic acid sequence extending from A_{24} to G_{44}, G_{159} to C_{182}, G_{561} to A_{596}, or A_{1011} to T_{1046} of SEQ ID NO:17.

The invention further provides a polynucleotide sequence comprising the complement, or fragments thereof, of any one of the polynucleotide sequences encoding SP. In another aspect, the invention provides compositions comprising isolated and purified polynucleotide sequences comprising the complements of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17, or fragments thereof.

The present invention further provides an expression vector containing at least a fragment of any one of the polynucleotide sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17. In yet another aspect, the expression vector containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide or a fragment thereof, the method comprising the steps of: a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding an SP under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified SP in conjunction with a suitable pharmaceutical carrier.

The invention also provides a purified antagonist of SP. In one aspect the invention provides a purified antibody which binds to an SP.

Still further, the invention provides a purified agonist of SP.

The invention also provides a method for treating or preventing a cancer, the method comprising the step of administering to a subject in need of such treatment an effective amount of a pharmaceutical composition containing SP.

The invention also provides a method for treating or preventing a cancer, the method

comprising the step of administering to a subject in need of such treatment an effective amount of an antagonist of SP.

The invention also provides a method for treating or preventing a neuronal disorder, the method comprising the step of administering to a subject in need of such treatment an effective amount of an antagonist of SP.

The invention also provides a method for treating or preventing an immune response associated with the increased expression or activity of SP, the method comprising the step of administering to a subject in need of such treatment an effective amount of an antagonist of SP.

The invention also provides a method for stimulating cell proliferation, the method comprising the step of administering to a cell an effective amount of purified SP.

The invention also provides a method for detecting a nucleic acid sequence which encodes a signal peptide-containing protein in a biological sample, the method comprising the steps of: a) hybridizing a nucleic acid sequence of the biological sample to a polynucleotide sequence complementary to the polynucleotide encoding SP, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the nucleic acid sequence encoding the signal peptide-containing protein in the biological sample.

The invention also provides a microarray which contains at least a fragment of at least one of the polynucleotide sequences encoding SP. In a particular aspect, the microarray contains at least a fragment of at least one of the sequences selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

The invention also provides a method for detecting the expression level of a nucleic acid sequence encoding a signal peptide-containing protein in a biological sample, the method comprising the steps of hybridizing the nucleic acid sequence of the biological sample to a complementary polynucleotide, thereby forming hybridization complex; and determining expression of the nucleic acid sequence encoding a signal peptide-containing protein in the biological sample by identifying the presence of the hybridization complex. In a preferred embodiment, prior to the hybridizing step, the nucleic acid sequences of the biological sample are amplified and labeled by the polymerase chain reaction.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A, 1B, 1C, 1D, and 1E show the amino acid sequence (SEQ ID NO:16) and nucleic acid sequence (SEQ ID NO:17) of SP16. The alignment was produced using MacDNASIS PRO™ software (Hitachi Software Engineering Co. Ltd. San Bruno, CA).

Figure 2 shows the amino acid sequence alignment between SP-16 (2547002; SEQ ID NO:16) and the bovine GPCR (GI 399711; SEQ ID NO:18) produced using the multisequence alignment program of DNASTAR™ software (DNASTAR Inc, Madison WI).

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, arrays and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

SP, as used herein, refers to the amino acid sequences of substantially purified SP

obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist", as used herein, refers to a molecule which, when bound to SP, increases or prolongs the duration of the effect of SP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of SP.

An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding SP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding SP as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent SP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding SP. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent SP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of SP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules. Fragments of SP are preferably about 5 to about 15 amino acids in length and retain the

biological activity or the immunological activity of SP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "antagonist" as used herein, refers to a molecule which, when bound to SP, decreases the amount or the duration of the effect of the biological or immunological activity of SP. Antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules which decrease the effect of SP.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, $F(ab')_2$, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind SP polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules include peptide nucleic acids and may be produced by

any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic SP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

A "composition comprising a given polynucleotide sequence" as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding SP (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17) or fragments thereof may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended using XL-PCR™ (Perkin Elmer,

Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly (e.g., GELVIEW™ Fragment Assembly system, GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of a ribonucleic acid that is similar to a polynucleotide encoding an SP by northern analysis is indicative of the presence of mRNA encoding SP in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

The term "SP" refers to any or all of the human polypeptides, SP-1, SP-2, SP-3, SP-4, SP-5, SP-6, SP-7, SP-8, SP-9, SP-10, SP-11, SP-12, SP-13, SP-14, SP-15, and SP-16.

A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to SP or the encoded SP. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a

second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

Human artificial chromosomes (HACs) are linear microchromosomes which may contain DNA sequences of 10K to 10M in size and contain all of the elements required for stable mitotic chromosome segregation and maintenance (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

"Inflammation" as used herein is interchangeable with "immune response", both terms refer to a condition associated with trauma, immune disorders, and infectious or genetic diseases and are characterized by production of cytokines, chemokines, and other signaling molecules which activate cellular and systemic defense systems.

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

"Microarray" refers to an array of distinct oligonucleotides arranged on a substrate, such as paper, nylon or other type of membrane, filter, gel, polymer, chip, glass slide, or any other suitable support.

The term "modulate", as used herein, refers to a change in the activity of SP. For

example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of SP.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" are those nucleic acid sequences which are greater than 60 nucleotides than in length, and most preferably includes fragments that are at least 100 nucleotides or at least 1000 nucleotides, and at least 10,000 nucleotides in length.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or hybridization assays. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers","primers", "oligomers", and "probes", as commonly defined in the art.

"Peptide nucleic acid", PNA as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of an SP encompasses the full-length SP and fragments thereof.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding SP, or fragments thereof, or SP itself may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA (in solution or bound to a solid support, a tissue, a tissue print, and the like.

The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The interaction is dependent upon the presence of a particular structure (i.e., the antigenic

determinant or epitope) of the protein recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The terms "stringent conditions"or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of SP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of signal peptide-containing proteins, collectively referred to as SP and individually as SP-1, SP2, SP-3, Sp-4, SP-5, SP-6, SP-7, SP-8, SP-9, SP-10, SP-11, SP-12, SP-13, SP-14, SP-15, and SP-16, the polynucleotides encoding SP (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17), and the use of these compositions for the diagnosis, treatment or prevention of diseases associated with cell proliferation and cell signaling. Table 1 shows the sequence identification numbers, reference, Incyte Clone number, cDNA library, NCBI sequence identifier and GenBank description for each of the signal peptide-containing proteins disclosed herein.

SP-1 was identified in Incyte Clone 1221102 from the NEUTGMT01 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:1, derived from Incyte Clone 1221102 encodes a GPCR with homology to GI 1575512, the GPR19 gene. Electronic northern analysis showed the expression of this sequence in neuronal tissues and in stimulated granulocytes.

SP-2 was identified in Incyte Clone 1457779 from the COLNFET02 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:2, derived from Incyte Clone 1457779 encodes an ATP diphosphohydrolase with homology to GI 1842120. Electronic northern analysis showed the expression of this sequence in fetal colon.

SP-3 was identified in Incyte Clone 1682433 from the PROSNOT15 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:3, derived from Incyte Clone 1682433 encodes a signal peptide-containing protein with homology to GI 1070391, a transmembrane protein. Electronic northern analysis showed the expression of this sequence in fetal, cancerous or inflamed cells and tissues. In particular, it was associated with cancerous prostate, asthmatic lung, promonocytes and IL-5 stimulated mononuclear cells.

SP-4 was identified in Incyte Clone 1899132 from the BLADTUT06 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:4, derived from Incyte Clone 1899132 encodes a signal peptide containing protein with homology to GI 887602, a Saccharomyces cerevisiae protein. Electronic northern analysis showed the expression of this sequence in inflamed cells and tissues (62%) and cancerous tissues (25%). In particular, it was associated with stimulated promonocyte and mononuclear cells.

SP-5 was identified in Incyte Clone 1907344 from the CONNTUT01 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:5, derived from Incyte Clone 1907344 encodes a signal peptide containing protein with homology to GI 33715, immunoglobulin light chain. Electronic northern analysis showed the expression of this sequence in cancerous tissues (66%), fetal or infant cells and tissues (22%).

SP-6 was identified in Incyte Clone 1963651 from the BRSTNOT04 cDNA library

using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:6, derived from Incyte Clone 1963651 encodes a GPCR with homology to GI 1657623, orphan receptor RDC1. Electronic northern analysis showed the expression of this sequence only in BRSTNOT04, tissue associated with a ductal carcinoma removed during mastectomy.

SP-7 was identified in Incyte Clone 1976095 from the PANCTUT02 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:7, derived from Incyte Clone 1976095 encodes a signal peptide-containing protein with homology to GI 2117185, a Mycobacterium tuberculosis protein. Electronic northern analysis showed the expression of this sequence in cancerous (50%) and inflamed (30%) tissues.

SP-8 was identified in Incyte Clone 2417676 from the HNT3AZT01 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:8, derived from Incyte Clone 2417676 encodes a signal peptide-containing protein with homology to GI 2150012, a human transmembrane protein. Electronic northern analysis showed this sequence to be expressed widely in proliferating, cancerous or inflamed tissues.

SP-9 was identified in Incyte Clone 1805538 from the SINTNOT13 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:9, derived from Incyte Clone 1805538 encodes a signal peptide-containing protein with homology to GI 294502, an extracellular matrix protein. Electronic northern analysis showed this sequence to be expressed in inflamed tissues (87%).

SP-10 was identified in Incyte Clone 1869688 from the SKINBIT01 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:10, derived from Incyte Clone 1869688 encodes a signal peptide-containing protein with homology to GI 1562, a G3 serine/threonine kinase. Electronic northern analysis showed this sequence to be expressed widely in proliferating fetal and inflamed tissues.

SP-11 was identified in Incyte Clone 1880692 from the LEUKNOT03 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:11, derived from Incyte Clone 1880692 encodes a signal peptide-containing protein with homology to GI 1487910, a Caenorhabditis elegans protein. Electronic northern analysis showed this sequence to be expressed in cancer and blood cells.

SP-12 was identified in Incyte Clone 318060 from the EOSIHET02 cDNA library

using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:12, derived from Incyte Clone 318060 encodes a receptor with homology to GI 606788, an opioid GPCR. Electronic northern analysis showed this sequence to be expressed in inflamed nerve and blood cells.

SP-13 was identified in Incyte Clone 396450 from the PITUNOT02 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:13, derived from Incyte Clone 396450 encodes a signal peptide-containing protein with homology to GI 342279, opiomelanocortin. Electronic northern analysis showed this sequence to be expressed in hormone producing cells and tissues (78%) and inflamed cells and tissues (45%).

SP-14 was identified in Incyte Clone 506333 from the TMLR3DT02 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:14, derived from Incyte Clone 506333 encodes a signal peptide-containing protein with homology to GI 2204110, adenylyl cyclase. Electronic northern analysis showed this sequence to be expressed widely in cancerous and inflamed cells and tissues.

SP-15 was identified in Incyte Clone 764465 from the LUNGNOT04 cDNA library using a computer search for amino acid sequence alignments. A nucleotide sequence, SEQ ID NO:15, derived from Incyte Clone 764465 encodes a receptor with homology to GI 1902984, lectin-like oxidized LDL receptor. Electronic northern analysis showed this sequence to be expressed in lung and in fetal liver .

SP-16 (SEQ ID NO:16) was identified in Incyte Clone 2547002 from the UTRSNOT11 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:17, was derived from the extension and assembly of the overlapping nucleic acid sequences of Incyte Clones 2741185 (BRSTTUT14), 2547002 (UTRSNOT11), and shotgun sequences, SAEA01463, SAEA01125, and SAEA00333.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:16, as shown in Figure 1A, 1B, 1C, 1D, and 1E. SP-16 is 350 amino acids in length and has a G protein coupled receptor signature at S_{125}GMQFLACISIDRYVAV; three potential N-glycosylation sites at N_6, N_{19}, and N_{276}; a potential glycosaminoglycan attachment site at S_{148}; and ten potential phosphorylation sites at S_{25}, T_{74}, T_{177}, S_{195}, T_{223}, Y_{269}, S_{278}, S_{309}, S_{323}, and S_{330}. SP-16 has 86% sequence identity with a bovine GPCR (GI 399711) and shares the GPCR signature, the N-glycosylation, the

glycosaminoglycan attachment site, and the first nine of the phosphorylation sites with the bovine receptor (Figure 2). Fragments of the nucleic acid sequence useful for designing oligonucleotides or to be used directly as hybridization probes to distinguish between these homologous molecules include A_{24} to G_{44}, G_{159} to C_{182}, G_{561} to A_{596}, or A_{1011} to T_{1046}. mRNA encoding SP-16 was expressed in cDNA libraries with inflamed smooth muscle cells, uterus (38%) and heart and blood vessel (38%).

The invention also encompasses SP variants which retain the biological or functional activity of SP. A preferred SP variant is one having at least 80%, and more preferably 90%, amino acid sequence identity to the SP amino acid sequence. A most preferred SP variant is one having at least 95% amino acid sequence identity to an SP disclosed herein.

The invention also encompasses polynucleotides which encode SP. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of SP can be used to produce recombinant molecules which express SP. In a particular embodiment, the invention encompasses a polynucleotide consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding SP, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring SP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode SP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring SP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding SP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which

particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof, which encode SP and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding SP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17, under various conditions of stringency as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by GIBCO/BRL (Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding SP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and

a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated,

and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode SP may be used in recombinant DNA molecules to direct expression of SP, fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express SP.

As will be understood by those of skill in the art, it may be advantageous to produce SP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter SP encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding SP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of SP activity, it may be useful to encode a chimeric SP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the SP encoding sequence and the heterologous protein sequence, so that SP may be cleaved and

purified away from the heterologous moiety.

In another embodiment, sequences encoding SP may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of SP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of SP, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active SP, the nucleotide sequences encoding SP or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding SP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding SP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with

virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid (GIBCO/BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding SP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for SP. For example, when large quantities of SP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding SP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or

factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding SP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express SP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding SP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of SP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which SP may be expressed (Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding SP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing SP in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition,

transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding SP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding SP, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express SP may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.

Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being used widely not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding SP is inserted within a marker gene sequence, transformed cells containing sequences encoding SP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding SP and express SP may be identified by a variety of procedures known to those of skill in the art.

These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding SP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding SP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding SP to detect transformants containing DNA or RNA encoding SP.

A variety of protocols for detecting and measuring the expression of SP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SP is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SP include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding SP, or any fragments thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding SP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly

depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode SP may be designed to contain signal sequences which direct secretion of SP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding SP to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and SP may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing SP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying SP from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

In addition to recombinant production, fragments of SP may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of SP may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists among the signal peptide-containing proteins of the invention. The expression of SP is closely associated with cell proliferation and cell signaling. Therefore, in atherosclerosis, cancers, immune response, or neuronal disorders where SP is an activator, hormone, transcription factor, or any other signaling molecule which promotes cell proliferation or signaling; it is desirable to decrease the

expression of SP. In cancers where SP is an inhibitor or suppressor and is controlling or decreasing cell proliferation, it is desirable to provide the protein or to increase the expression of SP.

In one embodiment, where SP is an inhibitor, SP or a fragment or derivative thereof may be administered to a subject to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a pharmaceutical composition comprising purified SP may be used to treat or prevent a cancer including, but not limited to, those listed above.

In another embodiment, an agonist which is specific for SP may be administered to a subject to treat or prevent a cancer including, but not limited to, those listed above.

In another further embodiment, a vector capable of expressing SP, or a fragment or a derivative thereof, may be administered to a subject to treat or prevent a cancer including, but not limited to, those listed above.

In a further embodiment where SP is promoting cell proliferation, antagonists which decrease the expression or activity of SP may be administered to a subject to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind SP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express SP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding SP may be administered to a subject to treat or prevent a cancer including, but not limited to, those listed above.

In one embodiment, where SP is an activator or stimulates cell signaling, an antagonist of SP may be administered to a subject to treat or prevent a neuronal disorder. Such disorders may be include, but are not limited to akathesia, Alzheimer's disease,

amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's disorder.

In another further embodiment, a vector expressing the complement of the polynucleotide encoding SP may be administered to a subject to treat or prevent a neuronal disorder, including, but not limited to, those listed above.

In yet another embodiment where SP is promoting cell proliferation, inflammation or immune response, antagonists which decrease the activity of SP may be administered to a subject to treat or prevent an immune response. Such responses may be associated with conditions and disorders such as atherosclerosis, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In particular, one aspect, antibodies which specifically bind SP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express SP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding SP may be administered to a subject to treat or prevent an immune response including, but not limited to, those associated with the disorders listed above

In one further embodiment, SP or a fragment or derivative thereof may be added to cells to stimulate cell proliferation. In particular, SP may be added to a cell in culture or cells in vivo using delivery mechanisms such as liposomes, viral based vectors, or electroinjection for the purpose of promoting cell proliferation and tissue or organ regeneration. Specifically, SP may be added to a cell, cell line, tissue or organ culture in vitro or ex vivo to stimulate cell proliferation for use in heterologous or autologous transplantation. In some cases, the cell will have been preselected for its ability to fight an infection or a cancer or to correct a genetic defect in a disease such as sickle cell anemia, β thalassemia, cystic fibrosis, or

Huntington's chorea.

In another embodiment, an agonist which is specific for SP may be administered to a cell to stimulate cell proliferation, as described above.

In another embodiment, a vector capable of expressing SP, or a fragment or a derivative thereof, may be administered to a cell to stimulate cell proliferation, as described above.

In other embodiments, any of the therapeutic proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of SP may be produced using methods which are generally known in the art. In particular, purified SP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SP.

Antibodies to SP may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with SP or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SP have an amino acid sequence consisting of at least five amino acids and more

preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of SP amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to SP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for SP may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between SP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SP epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding SP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding SP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding SP. Thus, complementary molecules or fragments may be used to modulate SP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding SP.

Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding SP. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra).

Genes encoding SP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes SP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions of the gene encoding SP (signal sequence, promoters, enhancers, and

introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or chaperons. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding SP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding SP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life.

Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino polymers (Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of SP, antibodies to SP, mimetics, agonists, antagonists, or inhibitors of SP. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used

pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions

may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of SP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SP or fragments thereof, antibodies of SP, agonists, antagonists or inhibitors of SP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50.

Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind SP may be used for the diagnosis of conditions or diseases characterized by expression of SP, or in assays to monitor patients being treated with SP, agonists, antagonists or inhibitors. The antibodies useful for

diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for SP include methods which utilize the antibody and a label to detect SP in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring SP are known in the art and provide a basis for diagnosing altered or abnormal levels of SP expression. Normal or standard values for SP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to SP under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, but preferably by photometric, means. Quantities of SP expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding SP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of SP may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of SP, and to monitor regulation of SP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding SP or closely related molecules, may be used to identify nucleic acid sequences which encode SP. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding SP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the SP encoding sequences. The

hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17, or fragments encompassing the nucleic acid sequence A_{24} to G_{44}, G_{159} to C_{182}, G_{561} to A_{596}, or A_{1011} to T_{1046} of SEQ ID NO:17, or from genomic sequences including promoter, enhancer elements, and introns of the naturally occurring SP.

Means for producing specific hybridization probes for DNAs encoding SP include the cloning of nucleic acid sequences encoding SP or SP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding SP may be used for the diagnosis of conditions, disorders, or diseases which are associated with either increased or decreased expression of SP. Examples of such conditions, disorders or diseases include cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers of the adrenal gland, bladder, bone, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, bone marrow, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; neuronal disorders such as akathesia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's disorder; and immune response associated with disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis,

pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and thyroiditis. The polynucleotide sequences encoding SP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays or microarrays utilizing fluids or tissues from patient biopsies to detect altered SP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding SP may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding SP may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding SP in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of SP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes SP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging

from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SP may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'->3') and another with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of SP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of disease, to diagnose disease, and to develop and monitor the activities of therapeutic agents.

In one embodiment, the microarray is prepared and used according to the methods known in the art such as those described in PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc.

Natl. Acad. Sci. 93: 10614-10619).

The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably about 15 to 30 nucleotides in length, and most preferably about 20 to 25 nucleotides in length. For a certain type of microarray, it may be preferable to use oligonucleotides which are only 7 to 10 nucleotides in length. The microarray may contain oligonucleotides which cover the known 5' (or 3') sequence, or may contain sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell or tissue type or to a normal, developmental, or disease state. In certain situations, it may be appropriate to use pairs of oligonucleotides on a microarray. The pairs will be identical, except for one nucleotide preferably located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from 2 to 1,000,000.

In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' or more preferably at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In one aspect, the oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

In one aspect, the oligonucleotides may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, such as that described in PCT application WO95/251116 (Baldeschweiler et al.). In another aspect, a "gridded" array analogous to a dot or slot blot (HYBRIDOT® apparatus, GIBCO/BRL) may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. In yet another aspect, an array may be produced by hand or by using available devices, materials,

and machines (including Brinkmann® multichannel pipettors or robotic instruments) and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other multiple from 2 to 1,000,000 which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarrays, polynucleotides are extracted from a biological sample. The biological samples may be obtained from any bodily fluid (blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. To produce probes, the polynucleotides extracted from the sample are used to produce nucleic acid sequences which are complementary to the nucleic acids on the microarray. If the microarray consists of cDNAs, antisense RNAs (aRNA) are appropriate probes. Therefore, in one aspect, mRNA is used to produce cDNA which, in turn and in the presence of fluorescent nucleotides, is used to produce fragment or oligonucleotide aRNA probes. These fluorescently labeled probes are incubated with the microarray so that the probe sequences hybridize to the cDNA oligonucleotides of the microarray. In another aspect, nucleic acid sequences used as probes can include polynucleotides, fragments, and complementary or antisense sequences produced using restriction enzymes, PCR technologies, and Oligolabeling or TransProbe kits (Pharmacia) well known in the area of hybridization technology.

Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies or functional analysis of the sequences, mutations, variants, or polymorphisms among samples (Heller, R.A. et al., (1997) Proc. Natl. Acad. Sci. 94:2150-55).

In another embodiment of the invention, the nucleic acid sequences which encode SP may be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries (cf.

Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154).

Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding SP on a physical chromosomal map and a specific disease , or predisposition to a specific disease, may help delimit the region of DNA associated with that disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, linkage analysis using established chromosomal markers, may be used to extend genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, and affected individuals.

In another embodiment of the invention, SP, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between SP and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to SP large numbers of different small test compounds are synthesized on a solid substrate, such as

plastic pins or some other surface. The test compounds are reacted with SP, or fragments thereof, and washed. Bound SP is then detected by methods well known in the art. Purified SP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding SP specifically compete with a test compound for binding SP. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with SP.

In additional embodiments, the nucleotide sequences which encode SP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

For purposes of example, the preparation and sequencing of the UTRSNOT11 cDNA library, from which Incyte Clone 2547002 was isolated, is described. Preparation and sequencing of cDNAs in libraries in the LIFESEQ™ database have varied over time, and the gradual changes involved use of kits, plasmids, and machinery available at the particular time the library was made and analyzed.

I UTRSNOT011 cDNA Library Construction

The UTRSNOT11 cDNA library was constructed from microscopically normal uterine tissue obtained from a 43-year-old female during a vaginal hysterectomy following the diagnosis of uterine leiomyoma. Pathology indicated that the myometrium contained an intramural leiomyoma and a submucosal leiomyoma. The endometrium was proliferative, however, the cervix and fallopian tubes were unremarkable. The right and left ovaries contained corpus lutea. The patient presented with metrorrhagia and deficiency anemia. Patient history included benign hypertension and atherosclerosis. Medications included Provera® tablets (medroxyprogesterone acetate; The Upjohn Company, Kalamazoo, MI),

iron and vitamins. Family history included benign hypertension in the father, atherosclerosis in a grandparent, malignant colon neoplasms in the mother, father, and a grandparent.

For the UTRSNOT11 library, the frozen tissue was homogenized and lysed in Trizol reagent (1 gm tissue/10 ml Trizol; Cat. #10296-028; GIBCO/BRL), a monoplastic solution of phenol and guanidine isothiocyanate, using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NY). After a brief incubation on ice, chloroform was added (1:5 v/v) and the lysate was centrifuged. The upper chloroform layer was removed to a fresh tube and the RNA extracted with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37°C. The RNA was re-extracted three times with acid phenol-chloroform pH 4.7 and precipitated using 0.3M sodium acetate and 2.5 volumes ethanol. The mRNA was isolated with the Qiagen Oligotex kit (QIAGEN, Inc., Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Cat. #18248-013, GIBCO/BRL). The cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into pINCY 1. The plasmid pINCY 1 was subsequently transformed into DH5a™ competent cells (Cat. #18258-012; GIBCO/BRL).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173, QIAGEN, Inc.). This kit enabled the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, GIBCO/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger, et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

III Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol 36:290-300; Altschul, et al. (1990) J. Mol. Biol. 215:403-410).

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms such as the one described in Smith, T. et al. (1992, Protein Engineering 5:35-51), incorporated herein by reference, could have been used when dealing with primary sequence patterns and secondary structure gap penalties. The sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-10} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam); and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp) for homology.

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., supra).

Analogous computer techniques use BLAST to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte

Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding SP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V Extension of SP Encoding Polynucleotides

The nucleic acid sequence of one of the nucleotide sequences of the present invention was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension in the antisense direction, and the other was synthesized to extend sequence in the sense direction. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be about 22 to about 30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures of about 68°to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension was necessary or desired, additional sets of primers were designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook et al., supra). After incubation for one hour at 37° C, the E. coli mixture was plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4

units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of one of the nucleotide sequences of the present invention were used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from one of the nucleotide sequences of the present invention are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 superfine resin column (Pharmacia & Upjohn). A aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are

sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours, hybridization patterns are compared visually.

VII Microarrays

To produce oligonucleotides for a microarray, one of the nucleotide sequences of the present invention are examined using a computer algorithm which starts at the 3' end of the nucleotide sequence. The algorithm identified oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that would interfere with hybridization. The algorithm identifies approximately 20 sequence-specific oligonucleotides of 20 nucleotides in length (20-mers). A matched set of oligonucleotides are created in which one nucleotide in the center of each sequence is altered. This processis repeated for each gene in the microarray, and double sets of twenty 20 mers are synthesized and arranged on the surface of the silicon chip using a light-directed chemical process, such as that discussed in Chee, supra.

In the alternative, a chemical coupling procedure and an ink jet device are used to synthesize oligomers on the surface of a substrate (cf. Baldeschweiler, supra). In another alternative, a "gridded" array analogous to a dot (or slot) blot is used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. A typical array may be produced by hand or using available materials and machines and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots. After hybridization, the microarray is washed to remove nonhybridized probes, and a scanner is used to determine the levels and patterns of fluorescence. The scanned image is examined to determine degree of complementarity and the relative abundance/expression level of each oligonucleotide sequence in the microarray.

VIII Complementary Polynucleotides

Sequence complementary to the sequence encoding SP, or any part thereof, is used to detect, decrease, or inhibit expression of naturally occurring SP. Although use of oligonucleotides comprising from about 15 to about 30 base-pairs is described, essentially the same procedure is used with smaller or larger sequence fragments. Appropriate

oligonucleotides are designed using Oligo 4.06 software and the coding sequence of one of the nucleotide sequences of the present invention. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the transcript encoding SP.

IX Expression of SP

Expression of SP is accomplished by subcloning the cDNAs into appropriate vectors and transforming the vectors into host cells. In this case, the cloning vector is also used to express SP in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of SP into the bacterial growth media which can be used directly in the following assay for activity.

X Demonstration of SP Activity

Cell proliferation SP may be expressed in a mammalian cell line such as DLD-1 or HCT116 (ATCC; Bethesda, MD) by transforming the cells with a eukaryotic expression vector encoding SP. Eukaryotic expression vectors are commercially available and the techniques to introduce them into cells are well known to those skilled in the art. The effect of SP on cell morphology may be visualized by microscopy; the effect on cell growth may be determined by measuring cell doubling-time; and the effect on tumorigenicity may be assessed by the ability of transformed cells to grow in a soft agar growth assay (Groden, J. et al. (1995) Cancer Res. 55:1531-1539).

Receptor Sp such as those encoded by SEQ ID NOs:17, 15, 12, 6 and 1 may be expressed in heterologous expression systems and their biological activity tested utilizing the purinergic receptor system (P_{2U}) as published by Erb, et al. (1993; Proc. Natl. Acad. Sci.

90:10449-53). Because cultured K562 human leukemia cells lack P_{2U} receptors, they can be transfected with expression vectors containing either normal or chimeric P_{2U} and loaded with fura-a, fluorescent probe for Ca^{++}. Activation of properly assembled and functional extracellular SP-transmembrane/intracellular P_{2U} receptors with extracellular UTP or ATP mobilizes intracellular Ca^{++} which reacts with fura-a and is measured spectrofluorometrically. Bathing the transfected K562 cells in microwells containing appropriate ligands will trigger binding and fluorescent activity defining effectors of SP. Once ligand and function are established, the P_{2U} system is useful for defining antagonists or inhibitors which block binding and prevent such fluorescent reactions.

XI Production of SP Specific Antibodies

SP that is substantially purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from one of the nucleotide sequences of the present invention is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (supra), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al., supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio iodinated, goat anti-rabbit IgG.

XII Purification of Naturally Occurring SP Using Specific Antibodies

Naturally occurring or recombinant SP is substantially purified by immunoaffinity chromatography using antibodies specific for SP. An immunoaffinity column is constructed by covalently coupling SP antibody to an activated chromatographic resin, such as

CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing SP is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/protein binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and SP is collected.

XIII Identification of Molecules Which Interact with SP

SP or biologically active fragments thereof are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133: 529). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SP, washed and any wells with labeled SP complex are assayed. Data obtained using different concentrations of SP are used to calculate values for the number, affinity, and association of SP with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified signal peptide-containing protein (SP) comprising a polypeptide having an amino acid sequence encoded by the polynucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

2. An isolated and purified polynucleotide sequence which hybridizes to the polynucleotide sequence encoding an SP of claim 1.

3 A composition comprising the polynucleotide sequence of claim 2.

4. An isolated and purified polynucleotide sequence having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:17.

5. A microarray containing at least a fragment of at least one of the polynucleotides encoding an SP of claim 1.

6. The fragment of the polynucleotide sequence of SEQ ID NO:17 of claim 4 wherein said fragment comprises the nucleic acid sequence extending from A_{24} to G_{44}, G_{159} to C_{182}, G_{561} to A_{596}, or A_{1011} to T_{1046}.

7. An isolated and purified polynucleotide having a nucleic acid sequence which is complementary to the nucleic acid sequence of the polynucleotide of claim 4.

8. A composition comprising the polynucleotide of claim 4.

9. An expression vector containing the polynucleotide of claim 4.

10. A host cell containing the vector of claim 9.

11. A method for producing a polypeptide encoding a signal peptide-containing protein, the method comprising the steps of:

a) culturing the host cell of claim 10 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

12. A pharmaceutical composition comprising a substantially purified signal peptide-containing a protein of claim 1 in conjunction with a suitable pharmaceutical carrier.

13. A purified antibody which binds specifically to the signal peptide-containing protein of claim 1.

14. A purified agonist which modulates the activity of the signal peptide-containing protein of claim 1.

15. A purified antagonist which decreases the effect of the signal peptide-containing protein of claim 1.

16. A method for stimulating cell proliferation, the method comprising administering to a cell an effective amount of the signal peptide-containing protein of claim 1.

17. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 12.

18. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 15.

19. A method for treating or preventing a neuronal disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 15.

20. A method for treating or preventing an immune response, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 15.

21. A method for detecting a nucleic acid sequence encoding a signal peptide-containing protein in a biological sample, the method comprising the steps of:

a) hybridizing the polynucleotide of claim 7 to the nucleic acid sequence of the biological sample, thereby forming a hybridization complex; and

b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the nucleic acid sequence encoding a signal peptide-containing protein in the biological sample.

22. A method for detecting the expression level of a nucleic acid sequence encoding a signal peptide-containing protein in a biological sample, the method comprising the steps of:

a) hybridizing the nucleic acid sequence of the biological sample to the

polynucleotides of claim 7, thereby forming a hybridization complex; and

b) determining expression of the nucleic acid sequence encoding the signal peptide-containing protein in the biological sample by identifying the presence of the hybridization complex.

23. The method of claim 22, wherein before hybridizating step, the polynucleotides of the biological sample are amplified and labeled by the polymerase chain reaction.

```
      9          18          27          36          45          54
NGC GAC GTA CAA CAG ATT GGA GCC ATG GCT TTG GAA CAG AAC CAG TCA ACA GAT
                                M   A   L   E   Q   N   Q   S   T   D

         63          72          81          90          99         108
TAT TAT TAT GAG GAA AAT GAA ATG AAT GGC ACT TAT GAC TAC AGT CAA TAT GAA
Y   Y   Y   E   E   N   E   M   N   G   T   Y   D   Y   S   Q   Y   E

        117         126         135         144         153         162
CTG ATC TGT ATC AAA GAA GAT GTC AGA GAA TTT GCA AAA GTT TTC CTC CCT GTA
L   I   C   I   K   E   D   V   R   E   F   A   K   V   F   L   P   V

        171         180         189         198         207         216
TTC CTC ACA ATA GTT TTC GTC ATT GGA CTT GCA GGC AAT TCC ATG GTA GTG GCA
F   L   T   I   V   F   V   I   G   L   A   G   N   S   M   V   V   A

        225         234         243         252         261         270
ATT TAT GCC TAT TAC AAG AAA CAG AGA ACC AAA ACA GAT GTG TAC ATC CTG AAT
I   Y   A   Y   Y   K   K   Q   R   T   K   T   D   V   Y   I   L   N

        279         288         297         306         315         324
TTG GCT GTA GCA GAT TTA CTC CTT CTA TTC ACT CTG CCT TTT TGG GCT GTT AAT
L   A   V   A   D   L   L   L   L   F   T   L   P   F   W   A   V   N
```

FIGURE 1A

```
          333         342         351         360         369         378
GCA GTT CAT GGG TGG GTT TTA GGG AAA ATA ATG TGC AAA ATA ACT TCA GCC TTG
A   V   H   G   W   V   L   G   K   I   M   C   K   I   T   S   A   L

          387         396         405         414         423         432
TAC ACA CTA AAC TTT GTC TCT GGA ATG CAG TTT CTG GCT TGT ATC AGC ATA GAC
Y   T   L   N   F   V   S   G   M   Q   F   L   A   C   I   S   I   D

          441         450         459         468         477         486
AGA TAT GTG GCA GTA ACT AAA GTC CCC AGC CAA TCA GGA GTG GGA AAA CCA TGC
R   Y   V   A   V   T   K   V   P   S   Q   S   G   V   G   K   P   C

          495         504         513         522         531         540
TGG ATC ATC TGT TTC TGT GTC TGG ATG GCT GCC ATC TTG CTG AGC ATA CCC CAG
W   I   I   C   F   C   V   W   M   A   A   I   L   L   S   I   P   Q

          549         558         567         576         585         594
CTG GTT TTT TAT ACA GTA AAT GAC AAT GCT AGG TGC ATT CCC ATT TTC CCC CGC
L   V   F   Y   T   V   N   D   N   A   R   C   I   P   I   F   P   R

          603         612         621         630         639         648
TAC CTA GGA ACA TCA ATG AAA GCA TTG ATT CAA ATG CTA GAG ATC TGC ATT GGA
Y   L   G   T   S   M   K   A   L   I   Q   M   L   E   I   C   I   G
```

FIGURE 1B

```
          657         666         675         684         693         702
TTT GTA GTA CCC TTT CTT ATT ATG GGG GTG TGC TAC TTT ATC ACA GCA AGG ACA
F   V   V   P   F   L   I   M   G   V   C   Y   F   I   T   A   R   T

          711         720         729         738         747         756
CTC ATG AAG ATG CCA AAC ATT AAA ATA TCT CGA CCC CTA AAA GTT CTG CTC ACA
L   M   K   M   P   N   I   K   I   S   R   P   L   K   V   L   L   T

          765         774         783         792         801         810
GTC GTT ATA GTT TTC ATT GTC ACT CAA CTG CCT TAT AAC ATT GTC AAG TTC TGC
V   V   I   V   F   I   V   T   Q   L   P   Y   N   I   V   K   F   C

          819         828         837         846         855         864
CGA GCC ATA GAC ATC ATC TAC TCC CTG ATC ACC AGC TGC AAC ATG AGC AAA CGC
R   A   I   D   I   I   Y   S   L   I   T   S   C   N   M   S   K   R

          873         882         891         900         909         918
ATG GAC ATC GCC ATC CAA GTC ACA GAA AGC ATC GCA CTC TTT CAC AGC TGC CTC
M   D   I   A   I   Q   V   T   E   S   I   A   L   F   H   S   C   L

          927         936         945         954         963         972
AAC CCA ATC CTT TAT GTT TTT ATG GGA GCA TCT TTC AAA AAC TAC GTT ATG AAA
N   P   I   L   Y   V   F   M   G   A   S   F   K   N   Y   V   M   K
```

FIGURE 1C

```
         981          990          999         1008         1017         1026
GTG GCC AAG AAA TAT GGG TCC TGG AGA AGA CAG AGA CAA AGT GTG GAG GAG TTT
V   A   K   K   Y   G   S   W   R   R   Q   R   Q   S   V   E   E   F

        1035         1044         1053         1062         1071         1080
CCT TTT GAT TCT GAG GGT CCT ACA GAG CCA ACC AGT ACT TTT AGC ATT TAA AGG
P   F   D   S   E   G   P   T   E   P   T   S   T   F   S   I

        1089         1098         1107         1116         1125         1134
TAA AAC TGC TCT GCC TTT TGC TTG GAT ACA TAT GAA TGA TGC TTT CCC CTC AAA

        1143         1152         1161         1170         1179         1188
TAA AAC ATC TGC ATT ATT CTG AAA CTC AAA TCT CAG ACG CCG TGG TTG CAA CTT

        1197         1206         1215         1224         1233         1242
ATA ATA AAG AAT GGG TTG GGG GAA GGG GGA GAA ATA AAA GCC AAG AAG AGG AAA

        1251         1260         1269         1278         1287         1296
CAA GAT AAT AAA TGT ACA AAA CAT GAA AAT TAA AAT GAA CAA TAT AGG AAA ATA

        1305         1314         1323         1332         1341         1350
ATT GTA ACA GGC ATA AGT GAA TAA CAC TCT GCT GTA ACG AAG AAG AGC TTT GTG
```

FIGURE 1D

```
      1359        1368        1377        1386        1395        1404
GTG ATA ATT TTG TAT CTT GGT TGC AGT GGT GCT TAT ACA AAT CTA CAC AAG TGA

      1413        1422        1431        1440        1449        1458
TAA AAT GAC AGA GAA CTA TAT ACA CAC ATT GTA CCA ATT TCA ATT TCC TGG TTT

      1467        1476        1485        1494        1503        1512
TGA CAT TAT AGT ATA ATT ATG TAA GAT GGA ACC ATT GGG GAA AAC TGG GTG AAG

      1521        1530        1539        1548        1557        1566
GGT ACC CAG GAC CAC TCT GTA CCA TCT TTG TAA CTT CCT GTG AAT TTA TAA TAA

      1575        1584        1593        1602        1611        1620
TTT CAA AAT AAA ACA AGT TAA AAA AAA AAC CCA CTA TGC TAT AAG TTA GGC CAT

      1629        1638        1647        1656
CTA AAA CAG ATT ATT AAA GAG GTT CAT GTT AAA AGG CAT GC
```

FIGURE 1E

```
1   M A L E Q N Q S T D Y Y Y E E N E M N G T Y D Y S Q Y E L I C I K E D V R E F A 2547002
1   M A V E Y N Q S T D Y Y Y E E N E M N D T H D Y S Q Y E V I C I K E E V R K F A GI 399711

41  K V F L P V F L T I V F V I G L A G N S M V V A I Y A Y Y K K Q R T K T D V Y I 2547002
41  K V F L P A F F T I A F I I G L A G N S T V V A I Y A Y Y K K R R T K T D V Y I GI 399711

81  L N L A V A D L L L L F T L P F W A V N A V H G W V L G K I M C K I T S A L Y T 2547002
81  L N L A V A D L F L L F T L P F W A V N A V H G W V L G K I M C K V T S A L Y T GI 399711

121 L N F V S G M Q F L A C I S I D R Y V A V T K V P S Q S G V G K P C W I I C F C 2547002
121 V N F V S G M Q F L A C I S T D R Y W A V T K A P S Q S G V G K P C W V I C F C GI 399711

161 V W M A A I L L S I P Q L V F Y T V N D N A R C I P I F P R Y L G T S M K A L I 2547002
161 V W V A A I L L S I P Q L V F Y T V N H K A R C V P I F P Y H L G T S M K A S I GI 399711

201 Q M L E I C I G F V V P F L I M G V C Y F I T A R T L M K M P N I K I S R P L K 2547002
201 Q I L E I C I G F I I P F L I M A V C Y F I T A K T L I K M P N I K K S Q P L K GI 399711

241 V L L T V V I V F I V T Q L P Y N I V K F C R A I D I I Y S L I T S C N M S K R 2547002
241 V L F T V V I V F I V T Q L P Y N I V K F C Q A I D I I Y S L I T D C D M S K R GI 399711

281 M D I A I Q V T E S I A L F H S C L N P I L Y V F M G A S F K N Y V M K V A K K 2547002
281 M D V A I Q I T E S I A L F H S C L N P V L Y V F M G T S F K N Y I M K V A K K GI 399711

321 Y G S W R R Q R Q S V E E F P F D S E G P T E P T S T F S I                   2547002
321 Y G S W R R Q R Q N V E E I P F E S E D A T E P T S T F S I                   GI 399711
```

FIGURE 2

<110> INCYTE PHARMACEUTICALS, INC.
LAL, Preeti
AU-YOUNG, Janice
REDDY, Roopa
MURRY, Lynn E.
MATHUR, Preete

<120> SIGNAL PEPTIDE-CONTAINING PROTEINS

<130> PF-0424 PCT

<140> To Be Assigned
<141> Herewith

<150> 08/966,316
<151> 1997-11-07

<160> 18

<170> PERL PROGRAM

<210> 1
<211> 619
<212> DNA
<213> Homo sapiens

<220> -
<223> 1221102

<400> 1
ggacaatgaa cattgtccct cggacaaaag tgaaaactat caagatgttc ctcattttaa 60
atctgttgtt tttgctctcc tggctgcctt ttcatgtagc tcagctatgg cacccccatg 120
aacaagacta taagaaaagt tcccttgttt tcacagctat cacatggata tcctttagtt 180
cttcagcctc taaacctact ctgtattcaa tttataatgc caatttcgga gagggatgaa 240
agagactttt tgcatgtcct ctatgaaatg ttaccgaagc aatgcctata ctatcacaac 300
aagttcaagg atggccaaaa aaaactacgt tggcatttca gaaatccctt ccatggccaa 360
aactattacc caaagactcg atctatgact catttgacag agaagccaag gaaaaaaagc 420
ttgcttggcc cattaactca aatccaccaa atacttttgt ccaagttctc attctttcaa 480
ttgttatgca ccagagatta aaaagcttta actataaaaa cagaagctat ttacatattt 540
gttttcactc aactttccaa gggaaatgtt ttattttgta aaatgcattc atttgtttac 600
tgtaaaaaaa aaaaaaaaa 619

<210> 2
<211> 742
<212> DNA
<213> Homo sapiens

<220> -
<223> 1457779

<400> 2
cctggagcca ggtgcacagc gcatcgcccg aggctgtcac cgccctgccc cgcccacccc 60
agctgtcctg gacccagggg cagggagagg ctggacgcca ggtgcgcgga cacagaagcg 120
tctaagcaca gcttcctcct tgccgctccg ggaagtgggc agccagccca ggaaccagta 180
ccacctgcac catggggctg tcccggaagg agcaggtctt cttggccctg ctgggggcct 240
cgggggtctc aggcctcacg gcactcattc tcctcctggt ggaggccacc agcgtgctcc 300
tgcccacaga catcaagttt gggatcgtgt ttgatgcggg ctcctcccac acgtccctct 360
tcctgtatca gtggccggcg aacaaggaga atggcacggg tgtggtcagc caggccctgg 420
cctgccaggt ggaagggcct ggaatctcct cctacacttc taatgctgca caggctggtg 480
agagcctgca gggctgcttg gaggaggcgc tggtgctgat cccagaggcc cagcatcgga 540
aaacacccac gttcctgggg gccacggctg gcatgaggtt gctcagccgg aagaacagct 600
ctcagggcca gggacatctt tgcagcagtc acccaggtcc tggggccggt ctcccgtgga 660
cttttggggt gccgagctcc tggccgggca ggccgaagtg gcctttggtt ggatcactgt 720
caactacggc ttggggacgt tt 742

<210> 3
<211> 1141
<212> DNA
<213> Homo sapiens

<220> -
<223> 1682433

<400> 3

cgctgaaacc ctgggcggcg gcaagctgtg cgacctcttc tgcggccggc ctgggcaggt 60
gtcttcctcg agaggcaggc aggggatccc ggacccttat acaggatgct gtgttctttg 120
ctcctttgtg aatgtctgtt gctggtagct ggttatgctc atgatgatga ctggattgac 180
cccacagaca tgcttaacta tgatgctgct tcaggaacaa tgagaaaatc tcaggcaaaa 240
tatggtattt caggggaaaa ggatgtcagt cctgacttgt catgtgctga tgaaatatca 300
gaatgttatc acaaacttga ttctttaact tataagattg atgagtgtga aaagaaaaag 360
agggaagact atgaaagtca aagcaatcct gtttttagga gatacttaaa taagatttta 420
attgaagctg gaaagcttgg acttcctgat gaaaacaaag gcgatatgca ttatgatgct 480

gagattatcc ttaaaagaga aactttgtta gaaatacaga agtttctcaa tggagaagac 540
tggaaaccag gtgccttgga tgatgcacta agtgatattt taattaattt taagtttcat 600
gattttgaaa catggaagtg gcgattcgaa gattcctttg gagtggatcc atataatgtg 660
ttaatggtac ttctttgtct gctctgcatc gtggttttag tggctaccga gctgtggaca 720
tatgtacgtt ggtacactca gttgagacgt gttttaatca tcagctttct gttcagtttg 780
ggatggaatt ggatgtattt atataagcta gcttttgcac agcatcaggc tgaagtcgcc 840
aagatggagc cattaaacaa tgtgtgtgcc aaaaagatgg actggactgg aagtatctgg 900
gaatggttta gaagttcatg gacctataag gatgacccat gccaaaaata ctatgagctc 960
ttactagtca accctatttg gttggtccca ccaacaaagg cacttgcagt tacattcacc 1020
acatttgtaa cggagccatt gaagcatatt ggaaaaggaa ctggggaatt tattaaagca 1080
ctcatgaagg aaattccagc gctgcttcat cttccagtgc tgataattat ggcattagcc 1140
a 1141

<210> 4
<211> 898
<212> DNA
<213> Homo sapiens

<220> -
<223> 1899132

<400> 4
tgcgaacctg gcccgtgcgg aaagggcgcg gagagccccg gcgcggagca ggcgggggac 60
ggtattcaga attcgagcgc aggagctccg cttctccacc tgctcccggg gagctattgg 120
gatccagaga atcacccgct gatggttttt gcccaggcct gaaacaacca gagagctacg 180
ggaaaggaag ggcttggctt gccagaggaa ttttccaagt gctcaaacgc caggcttacg 240
gcgcctgtga tccgtccagg aggacaaagt gggatttgaa gatccactcc acttctgctc 300
atggcgggcc agggcctgcc cctgcacgtg gccacactgc tgactgggct gctggaatgc 360
ctgggctttg ctggcgtcct ctttggctgg ccttcactag tgtttgtctt caagaatgaa 420
gattacttta aggatctgtg tggaccagat gctgggccga ttggcaatgc cacagggcag 480
gctgactgca aagcccagga tgagaggttc tcactcatct tcaccctggg gtccttcatg 540
aacaacttca tgacattccc cactggctac atctttgacc ggttcaagac caccgtggca 600
cgcctcatag ccatattttt ctacaccacc gccacactca tcatagcctt cacctctgca 660
ggctcagccg tgctgctctt cctggccatg ccaatgctca ccattggggg aatcctgttt 720
ctcatcacca acctgcagat tgggaaccta tttggccaac accgttcgac catcatcact 780
ctgtacaatg gagcatttga ctcttcctcg gcagtcttcc ttattattaa gcttctttat 840
gaaaaaggca tcagcctcag ggcctgcacc tggcgcctcg agcacgacta tatattgc 898

<210> 5
<211> 450

<212> DNA
<213> Homo sapiens

<220> -
<223> 1907344

<400> 5
gctcagctgt gggcttagga agcagagcct ggggcatctc caccatggcc tggacccctc 60
tcctcctcca gcttctcacc ctctgctcag ggtcctgggc acagtctgcg ctgacccagg 120
aagcctcggt gtcagggacc gtgggacaga aggtcaccct gtcctgttct ggaaacaaca 180
acaacattgg aagttatgct gtgggctggt accaacagat ttctcacggt gttctcaaaa 240
ctgtgatatt tggaaattct ccgccctcag ggatccctta ccgcttctct ggctcaaagt 300
ctgggaccac agcctccctg actatctcgg gcctccagcc tgaggacgag gctgattatt 360
atttttcaac atgggactac agactcagtg ctgtggtttt cggcggaagg accaaactga 420
ccgtcctagg tcagcccaag gctgcccccc 450

<210> 6
<211> 2111
<212> DNA
<213> Homo sapiens

<220> -
<223> 1963651

<400> 6

aagtgctcag cactaaggga gccagcgcac agcacagcca ggaaggcgag cgagcccagc 60
cagcccagcc agcccagcca gcccggaggt atctgtgaga taggtgctgc tgtcctgggg 120
aggtagatgc agacagatta actctcaagg tcatttgatt gcccgcctca gaacgatgga 180
tctgcatctc ttcgactact cagagccagg gaacttctcg gacatcagct ggccatgcaa 240
cagcagcgac tgcatcgtgg tggacacggt gatgtgtccc aacatgccca acaaaagcgt 300
cctgctctac acgctctcct tcatttacat tttcatcttc gtcatcggca tgattgccaa 360
ctccgtggtg gtctgggtga atatccaggc caagaccaca ggctatgaca cgcactgcta 420
catcttgaac ctggccattg ccgacctgtg ggttgtcctc accatyccag tctgggtggt 480
cagtctcgtg gmagcacaac cagtggccca tgggcgagct cacgtgcaaa gtcacacacc 540
tcatcttytc catcaacctc ttcggcagca ttttcttcct cacgtgcatg agcgtggacc 600
gctacctctc catcacctac ttcaccaaca cccccagcag caggaagaag atggtacgcc 660
gtgtcgtctg catcctggtg tggctgctgg ccttctgcgt gtctctgcct gacacctact 720
acctgaagac cgtcacgtct gcgtccaaca atgagaccta ctgccggtcc ttctaccccg 780
agcacagcat caaggagtgg ctgatcggca tggagctggt ctccgttgtc ttgggctttg 840
ccgttccctt ctccattatc gctgtcttct acttcctgct ggccagagcc atctcggcgt 900

ccagtgacca ggagaagcac agcagccgga agatcatctt ctcctacgtg gtggtcttcc 960
ttgtctgctg gttgccctac cacgtggcgg tgctgctgga catcttctcc atcctgcact 1020
acatcccttt cacctgccgg ctggagcacg ccctcttcac ggccctgcat gtcacacagt 1080
gcctgtcgct ggtgcactgc tgcgtcaacc ctgtcctcta cagcttcatc aatcgcaact 1140
acaggtacga gctgatgaag gccttcatct tcaagtactc ggccaaaaca gggctcacca 1200
agctcatcga tgcctccaga gtctcagaga cggagtactc tgccttggag cagagcacca 1260
aatgatctgc cctggagagg ctctgggacg ggtttacttg tttttgaaca gggtgatggg 1320
ccctatggtt ttctagrgca aagcaaagym scyycgggga aycyyratcc cccscttgag 1380
tccmsmgtga agaggggags acgtgcccca gcttggcatc cawtctctct tggkctcttg 1440
atgacgcagc tgtcatttgg ctgtaarcaa gtgctgacag ttttscaacr gggcagagct 1500
gttgtcscac agccagtgcc tgtgccgtca gagcccagct gaggacmggc ttgccckgga 1560
cctyctgawa agataggatt tyckgkgtty cckgaatttt twawatggkg attkgtattt 1620
aaawtttaag accttwattt ycycactatt ggkgkacctt ataaatgtat tkgaaagtta 1680
aatatatttt aaatattgtt tgggaggcat agtgctgaca tatattcaga gtgttgtagt 1740
tttaaggtta gcgtgacttc agttttgact aaggatgaca ctaattgtta gctgttttga 1800
aattatatat atataaatat atataaatat ataaatatat gccagtcttg gctgaaatgt 1860
tttatttacc atagttttat atctgtgtgg tgttttgtac cggcacggga tatggaacga 1920
aaactgcttt gtaatgcagt ttgtgacatt aatagtattg taaagttaca ttttaaaata 1980
aacaaaaaac tgttctggac tgcaaatctg cacacacaac gaacagttgc atttcagaga 2040
gttctctcaa tttgtaagtt attttttttt aataaagatt tttgtttcct aaaaatgcaa 2100
aaaaaaaaaa a 2111

<210> 7
<211> 700
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 21, 57
<223> a or g or c or t, unknown, or other

<220> -
<223> 1976095

<400> 7
gacgccagcg cctgcagagg ntgagcaggg aaaaagccag tgccccagcg gaagacnagc 60
tcagagctgg tctgccatgg acatcctggt cccactcctg cagctgctgg tgctgcttct 120
taccctgccc ctgcacctca tggctctgct gggctgctgg cagcccctgt gcaaaagcta 180
cttcccctac ctgatggccg tgctgactcc caagagcaac cgcaagatgg agagcaagaa 240
acgggagctc ttcagccaga taaaggggct tacaggagcc tccgggaaag tggccctact 300

ggagctgggc tgcggaaccg gagccaactt tcagttctac ccaccgggct gcagggtcac 360
ctgcctagac ccaaatcccc actttgagaa gttcctgaca aagagcatgg ctgagaacag 420
gcacctccaa tatgagcggt ttgtggtggc tcctggagag gacatgagac agctggctga 480
tggctccatg gatgtggtgg tctgcactct ggtgctgtgc tctgtgcaga gcccaaggaa 540
ggtcctgcag gaggtccgga gagtactgag accgggaggt gtgctctttt tctgggagca 600
tgtggcagaa ccatatggaa gctgggcctt catgtggcag caagttttcg agcccacctg 660
gaaacacatt ggggatggct tgctgcctca ccagagagac 700

<210> 8
<211> 363
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 330
<223> a or g or c or t, unknown, or other

<220> -
<223> 2417676

<400> 8
gggaatttcc cttatctcct tcgcagtgca gctccttcaa cctcgccatg gcctctgccg 60
gaatgcagat cctgggagtc gtcctgacac tgctgggctg ggtgaatggc ctggtctcct 120
gtgccctgcc catgtggaag gtgaccgctt tcatcggcaa cagcatcgtg gtggcccagg 180
tggtgtggga gggcctgtgg atgtcctgcg tggtgcagag caccggccag atgcagtgca 240
aggtgtacga ctcactgctg gcgctgccac aggacctgca ggctgcacgt gccctctgtg 300
tcatcgccct ccttgtggcc ctgttcggcn tgctggtcta ccttgctggg gccaagttta 360
cca 363

<210> 9
<211> 575
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 2, 4
<223> a or g or c or t, unknown, or other

<220> -
<223> 1805538

<400> 9
cngntcgagg ctaagaggac aggatgaggc ccggcctctc atttctccta gcccttctgt 60
tcttccttgg ccaagctgca ggggatttgg gggatgtggg acctccaatt cccagccccg 120
gcttcagctc tttcccaggt gttgactcca gctccagctt cagctccagc tccaggtcgg 180
gctccagctc cagccgcagc ttaggcagcg gaggttctgt gtcccagttg ttttccaatt 240
tcaccggctc cgtggatgac cgtgggacct gccagtgctc tgtttccctg ccagacacca 300
cctttcccgt ggacagagtg gaacgcttgg aattcacagc tcatgttctt tctcagaagt 360
ttgagaaaga actttccaaa gtgagggaat atgtccaatt aattagtgtg tatgaaaaga 420
aactgttaaa cctaatgtcc gaattgacat catggagaag gataccattt cttacactga 480
actggacttc gagctgatca aggtagaagt gaaggagatg gaaaaactgg tcatacagct 540
gaaggagagt ttggtggaag tcagaaattg ttgac 575

<210> 10
<211> 1637
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 3, 13, 22, 44, 69, 162, 1220, 1426, 1443, 1458, 1465, 1486,
<221> unsure
<222> 1488, 1490, 1517, 1522, 1524, 1525, 1533, 1553, 1573, 1584,
<221> unsure
<222> 1605, 1624, 1631, 1634, 1635
<223> a or g or c or t, unknown, or other

<220> -
<223> 1869688

<400> 10
acncagcctt ttncccgatt cnccctttcc tgccttcggt ttcntcccaa ttcttaccca 60
tcccctacna gctgccatcc ctgacaccct tctctcctgg gccacgcagt ccaacctgaa 120
cgggagcggg gaggtatcct ggcaccttcc ttggctctta cnctcggtt tctcacagcg 180
gggccggcgc cgccatggcg gccgtgtttg atttggattt ggagacggag gaaggcagcg 240
agggcgaggg cgagccagag ctcagccccg cggacgcatg tcccttgcc gagttgaggg 300
cagctggcct agagcctgtg ggacactatg aagaggtgga gctgactgag accagcgtga 360
acgttggccc agagcgcatc gggccccact gctttgagct gctgcgtgtg ctgggcaagg 420
ggggctatgg caaggtgttc caggtgcgaa aggtgcaagg caccaacttg ggcaaaatat 480

atgccatgaa agtcctaagg aaggccaaaa ttgtgcgcaa tgccaaggac acagcacaca 540
cacgggctga gcggaacatt ctagagtcag tgaagcaccc ctttattgtg gaactggcct 600
atgccttcca gactggtggc aaactctacc tcatccttgg attgcctcag tggtggcgag 660
ctcttcacgc atctggagcg agagggcatc ttcctggaag atacggcctg cttctacctg 720
gctgagatca cgctggccct gggccatctc cactcccagg gcatcatcta ccgggacctc 780
aagcccgaga acatcatgct cagcagccag ggccacatca aactgaccga ctttggactc 840
tgcaaggagt ctatccatga gggcgccgtc actcacacct tctgcggcac cattgagtac 900
atggcccctg agattctggt gcgcagtggc cacaaccggg ctgtggactg gtggagcctg 960
ggggccctga tgtacgacat gctcactgga tcgccgccct tcaccgcaga gaaccggaag 1020
aaaaccatgg ataagatcat caggggcaag ctggcactgc ccccctacct caccccagat 1080
gcccgggacc ttgtcaaaaa gtttctgaaa cggaatccca gccagcggat tgggggtggc 1140
ccaggggatg ctgctgatgt gcagagacat ccctttttcc ggcacatgaa ttgggacgac 1200
ttctggcctg gcgtgtggan ccccctttca aggccctgtc tgcagtcaga ggagacgtga 1260
gcagtttgat acccgcttca cacggcagac gccggtggac agtcctgatg acacagcctc 1320
agcgagagtg ccaacaaggc cttcctgggg ttacataagt ggcgcgtctg tcctggacag 1380
atcaagaggt tctctttcag cccaagtggg tcaaccaggg ctcaanatag ccccgggtcc 1440
gtnagcccct caagtttncc ctttnagggt tcggccagcc accttncngn gccaaggagt 1500
acttactcaa tctgcanggg gngnnttgac aangcctttt ccatcgtccc ctnagggcaa 1560
aattaaaagg gcntgggtta aggntagaac cggtggggta taagntccct tagccgtcct 1620
gggnttaaaa naanntg 1637

<210> 11
<211> 1124
<212> DNA
<213> Homo sapiens

<220> -
<223> 1880692

<400> 11
ggaagagcag cggcgaggcg gcggtggtgg ctgagtccgt ggtggcagag gcgaaggcga 60
cagctctagg ggttggcacc ggccccgaga ggaggatgcg ggtccggata gggctgacgc 120
tgctgctgtg tgcggtgctg ctgagcttgg cctcggcgtc ctcggatgaa gaaggcagcc 180
aggatgaatc cttagattcc aagactactt tgacatcaga tgagtcagta aaggaccata 240
ctactgcagg cagagtagtt gctggtcaaa tatttcttga ttcagaagaa tctgaattag 300
aatcctctat tcaagaagag gaagacagcc tcaagagcca agaggggaa agtgtcacag 360
aagatatcag ctttctagag tctccaaatc cagaaaacaa ggactatgaa gagccaaaga 420
aagtacggaa accagctttg accgccattg aaggcacagc acatggggag ccctgccact 480
tccctttct tttcctagat aaggagtatg atgaatgtac atcagatggg agggaagatg 540
gcagactgtg gtgtgctaca acctatgact acaaagcaga tgaaaagtgg ggcttttgtg 600
aaactgaaga agaggctgct aagagacggc agatgcagga agcagaaatg atgtatcaaa 660

ctggaacgaa aatccttaat ggaagcaata agaaaagcca aaaaagagaa gcatatcggt 720
atctccaaaa ggcagcaagc atgaaccata ccaaagccct ggagagagtg tcatatgctc 780
ttttatttgg tgattacttg ccacagaata tccaggcagc gagagagatg tttgagaagc 840
tgactgagga aggctctccc aagggacaga ctgctcttgg ctttctgtat gcctctggac 900
ttggtgttaa ttcaagtcag gcaaaggctc ttgtatatta tacatttgga gctcttgggg 960
gcaatctaat agcccacatg gttttgggtt acagatactg ggctggcatc ggcgtcctcc 1020
agagttgtga atctgccctg actcactatc gtcttgttgc caatcatggt atctatgttt 1080
ccccttttac cttttaggaa aaaaaaataa atggaattaa cttt 1124

<210> 12
<211> 1452
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 3, 472, 484, 486, 499, 501, 502, 504, 508, 513, 572, 577,
<221> unsure
<222> 637, 642, 646, 650, 655, 669, 688, 698
<223> a or g or c or t, unknown, or other

<220> -
<223> 318060

<400> 12
cancaggtgt ttattagggt ccttttcat taccccagag acagacccag ggctggctac 60
gtgcacagga agtaacgctt gccacatgca taaatacgtg aaggtgcaca ttacatcagc 120
acagattcac aaaacacctc gccttggcaa gaaaactgta gctaggcagc tcccgtcctc 180
agggactcct gccacagacg tcatggagac agcatgagcc tccccagaac agtccccacg 240
gcctagactc cccagagcag gaggagcagc ccaggctctg ttgcgagaca gccatcactt 300
cctgttcttt gcaggtgcct aaggtaggtt acctggccaa ggttttggtg gaaaaaatga 360
gtttttttcaa tgttgcaggt cttttaatag ttcatctgta ggaagtgcat ttgcaaagtc 420
accaacctgc agcttccatc tgtagaccag gaagggtgat tctctgggtg ancacagcgg 480
ggcntnccct gaggtacana nntncccncc canacccccg cagtgtcctc acagccatca 540
caggctttgg aagtttggct caagcaaggc cnttgcnaag gcccccaacc cccttcatgg 600
ttgggcttct gctgtgaaag ccaatccctc ccggttnggg cnagcnaagn tcaangggcc 660
ttaccccang aggccattct tgaagggntt gtaaaatnga agcaggaagc tgtgtggaag 720
gagaagctgg tggccacagc agagtcctgc tctggggacg cctgcttcat ttacaagcct 780
caagatggct ctgtgtaggg cctgagcttg ctgcccaacg ggaggatggc ttcacagcag 840
agccagcatg aggggtgggg cctggcaggg cttgcttgag ccaaactgca aaggctgtgg 900
tggctgtgag gacactgcgg gggttggggg ggggcgtctg tacctcaggg gatgccccgc 960
tgtggtcacc cagagaatca cccttcctgg tctacagatg gaagctgcag gttggtgact 1020

ttgcaaatgc acttcctaca gatgaactat taaaagacct gcaacattga aaaaactcat 1080
ttttccacc aaaaccttgg ccaggtaacc taccttaggc acctgcaaag aacaggaagt 1140
gatggctgtc tcgcaacaga gcctgggctg ctcctcctgc tctggggagt ctaggccgtg 1200
gggactgttc tggggaggct catgctgtct ccatgacgtc tgtggcagga gtccctgagg 1260
acgggagctg cctaagctac agtttttytt sccaagggcg aggtgttttg tgaatctgtg 1320
ctgatgtaat gtgcaccttc acgtatttat gcatgtggca agcgttactt cctgtgcacg 1380
tagccagccc tgggtctgtc tctggggtaa tgaaaaagga ccctaataaa cacctgctca 1440
ctggctgggt gg 1452

<210> 13
<211> 280
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 19, 29, 43, 49, 69, 75, 86, 112, 115, 130, 185, 200, 244,
<221> unsure
<222> 252, 254, 267, 278
<223> a or g or c or t, unknown, or other

<220> -
<223> 396450

<400> 13
ggggaagaag agccgcganc gagagaggnc ggcgagcgtc ccnggcctna gagagcagcc 60
tcccgagana ggcanttgct ggattntcca aaagtatctg cagtggctgt tncancagga 120
gagcctcagn ctgcctggaa gatgccgaga tcgtgctgca gccgctcggg ggccctgttg 180
ctggncttgc tgcttcaggn ctccatggaa gtgcgtggct ggtgcctgga gagcagccag 240
tgtnaggacc tnancaagga aagcaanctg cttgagtnca 280

<210> 14
<211> 514
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 378, 393, 428, 444, 460
<223> a or g or c or t, unknown, or other

<220> -
<223> 506333

<400> 14
tgtggagtca gcccagtctg gatgcacagg aggatgctgg cggcacagtg agtgaggcct 60
ggtgccagag ctgtgcggac cccttgttgg ccatggagca gcaggcccag aggccctctc 120
cccagccctg cttgcctgcc tcggagagga cagaggccta ggcccacggg ggagggtgtt 180
ggcagacaga tgccctccag gccctggggc ctccttaacg gcccccttaac gacacgcgtg 240
ccaagggtgg aggatgccag ccaaggggcg ctacttcctc aacgagggcg aggagggccc 300
tgaccaagat gcgctctacg agaagtacca gctcaccagc cagcatgggc cgctgctgct 360
cacgctcctg ctggtggncg caatgcctgc gtngccctca tcatattgcc tcagccaggg 420
ggtgagtnaa ggcagcccct gggntcaagt ctcggcccan actttggcaa gtgctatctt 480
ctcttagctc ttctgaaaat gcttatcttc tgta 514

<210> 15
<211> 617
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 537, 578, 598, 606
<223> a or g or c or t, unknown, or other

<220> -
<223> 764465

<400> 15
aaactacatt ttgcaaagtc attgaactct gagctcagtt gcagtactcg ggaagccatg 60
caggatgaag atggatacat caccttaaat attaaaactc ggaaaccagc tctcgtctcc 120
gttggccctg catcctcctc ctggtggcgt gtgatggctt tgattctgct gatcctgtgc 180
gtggggatgg ttgtcgggct ggtggctctg gggatttggt ctgtcatgca gcgcaattac 240
ctacaagatg agaatgaaaa tcgcacagga actctgcaac aattagcaaa gcgcttctgt 300
caatatgtgg taaaacaatc agaactaaaa gggcactttc aaaggtcata aatgcagccc 360
ctgtgacaca aactggagat attatggaga tagctgctat gggttcttca ggcacaactt 420
aacatgggaa gagagtaagc agtactgcac tgacatgaat gctactctcc tgaagattga 480
caaccggaac attgtggagt acatcaaagc caggactcat ttaattcgtt tgggtcngat 540
tatctcgcca gaagtcgaat gaggtctgga agtggganga tggctcgggt atctcagnaa 600
atatgnttga gtttttg 617

<210> 16
<211> 350
<212> PRT
<213> Homo sapiens

<220> -
<223> 2547002

<400> 16
Met Ala Leu Glu Gln Asn Gln Ser Thr Asp Tyr Tyr Tyr Glu Glu
1 5 10 15
Asn Glu Met Asn Gly Thr Tyr Asp Tyr Ser Gln Tyr Glu Leu Ile
20 25 30
Cys Ile Lys Glu Asp Val Arg Glu Phe Ala Lys Val Phe Leu Pro
35 40 45
Val Phe Leu Thr Ile Val Phe Val Ile Gly Leu Ala Gly Asn Ser
50 55 60
Met Val Val Ala Ile Tyr Ala Tyr Tyr Lys Lys Gln Arg Thr Lys
65 70 75
Thr Asp Val Tyr Ile Leu Asn Leu Ala Val Ala Asp Leu Leu Leu
80 85 90
Leu Phe Thr Leu Pro Phe Trp Ala Val Asn Ala Val His Gly Trp
95 100 105
Val Leu Gly Lys Ile Met Cys Lys Ile Thr Ser Ala Leu Tyr Thr
110 115 120
Leu Asn Phe Val Ser Gly Met Gln Phe Leu Ala Cys Ile Ser Ile
125 130 135
Asp Arg Tyr Val Ala Val Thr Lys Val Pro Ser Gln Ser Gly Val
140 145 150
Gly Lys Pro Cys Trp Ile Ile Cys Phe Cys Val Trp Met Ala Ala
155 160 165
Ile Leu Leu Ser Ile Pro Gln Leu Val Phe Tyr Thr Val Asn Asp
170 175 180
Asn Ala Arg Cys Ile Pro Ile Phe Pro Arg Tyr Leu Gly Thr Ser
185 190 195
Met Lys Ala Leu Ile Gln Met Leu Glu Ile Cys Ile Gly Phe Val
200 205 210
Val Pro Phe Leu Ile Met Gly Val Cys Tyr Phe Ile Thr Ala Arg
215 220 225
Thr Leu Met Lys Met Pro Asn Ile Lys Ile Ser Arg Pro Leu Lys
230 235 240
Val Leu Leu Thr Val Val Ile Val Phe Ile Val Thr Gln Leu Pro

245 250 255
Tyr Asn Ile Val Lys Phe Cys Arg Ala Ile Asp Ile Ile Tyr Ser
260 265 270
Leu Ile Thr Ser Cys Asn Met Ser Lys Arg Met Asp Ile Ala Ile
275 280 285
Gln Val Thr Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro
290 295 300
Ile Leu Tyr Val Phe Met Gly Ala Ser Phe Lys Asn Tyr Val Met
305 310 315
Lys Val Ala Lys Lys Tyr Gly Ser Trp Arg Arg Gln Arg Gln Ser
320 325 330
Val Glu Glu Phe Pro Phe Asp Ser Glu Gly Pro Thr Glu Pro Thr
335 340 345
Ser Thr Phe Ser Ile
350

<210> 17
<211> 1660
<212> DNA
<213> Homo sapiens

<220> -
<223> 2547002

<400> 17
gcgacgtaca acagattgga gccatggctt tggaacagaa ccagtcaaca gattattatt 60
atgaggaaaa tgaaatgaat ggcacttatg actacagtca atatgaactg atctgtatca 120
aagaagatgt cagagaattt gcaaaagttt tcctccctgt attcctcaca atagttttcg 180
tcattggact tgcaggcaat tccatggtag tggcaattta tgcctattac aagaaacaga 240
gaaccaaaac agatgtgtac atcctgaatt tggctgtagc agatttactc cttctattca 300
ctctgccttt ttgggctgtt aatgcagttc atgggtgggt tttagggaaa ataatgtgca 360
aaataacttc agccttgtac acactaaact ttgtctctgg aatgcagttt ctggcttgta 420
tcagcataga cagatatgtg gcagtaacta aagtccccag ccaatcagga gtgggaaaac 480
catgctggat catctgtttc tgtgtctgga tggctgccat cttgctgagc ataccccagc 540
tggtttttta tacagtaaat gacaatgcta ggtgcattcc cattttcccc cgctacctag 600
gaacatcaat gaaagcattg attcaaatgc tagagatctg cattggattt gtagtaccct 660
ttcttattat gggggtgtgc tactttatca cagcaaggac actcatgaag atgccaaaca 720
ttaaaatatc tcgacccctа aaagttctgc tcacagtcgt tatagttttc attgtcactc 780
aactgcctta taacattgtc aagttctgcc gagccataga catcatctac tccctgatca 840
ccagctgcaa catgagcaaa cgcatggaca tcgccatcca agtcacagaa agcatcgcac 900
tctttcacag ctgcctcaac ccaatccttt atgtttttat gggagcatct ttcaaaaact 960

acgttatgaa agtggccaag aaatatgggt cctggagaag acagagacaa agtgtggagg 1020
agtttccttt tgattctgag ggtcctacag agccaaccag tacttttagc atttaaaggt 1080
aaaactgctc tgcctttgc ttggatacat atgaatgatg ctttcccctc aaataaaaca 1140
tctgcattat tctgaaactc aaatctcaga cgccgtggtt gcaacttata ataaagaatg 1200
ggttggggga agggggagaa ataaaagcca agaagaggaa acaagataat aaatgtacaa 1260
aacatgaaaa ttaaaatgaa caatatagga aaataattgt aacaggcata agtgaataac 1320
actctgctgt aacgaagaag agctttgtgg tgataatttt gtatcttggt tgcagtggtg 1380
cttatacaaa tctacacaag tgataaaatg acagagaact atatacacac attgtaccaa 1440
tttcaatttc ctggttttga cattatagta taattatgta agatggaacc attggggaaa 1500
actgggtgaa gggtacccag gaccactctg taccatcttt gtaacttcct gtgaatttat 1560
aataatttca aaataaaaca agttaaaaaa aaaacccact atgctataag ttaggccatc 1620
taaaacagat tattaaagag gttcatgtta aaaggcatgc 1660

<210> 18
<211> 350
<212> PRT
<213> Bos taurus

<220> -
<223> g399711

<400> 18
Met Ala Val Glu Tyr Asn Gln Ser Thr Asp Tyr Tyr Tyr Glu Glu
1 5 10 15
Asn Glu Met Asn Asp Thr His Asp Tyr Ser Gln Tyr Glu Val Ile
20 25 30
Cys Ile Lys Glu Glu Val Arg Lys Phe Ala Lys Val Phe Leu Pro
35 40 45
Ala Phe Phe Thr Ile Ala Phe Ile Ile Gly Leu Ala Gly Asn Ser
50 55 60
Thr Val Val Ala Ile Tyr Ala Tyr Tyr Lys Lys Arg Arg Thr Lys
65 70 75
Thr Asp Val Tyr Ile Leu Asn Leu Ala Val Ala Asp Leu Phe Leu
80 85 90
Leu Phe Thr Leu Pro Phe Trp Ala Val Asn Ala Val His Gly Trp
95 100 105
Val Leu Gly Lys Ile Met Cys Lys Val Thr Ser Ala Leu Tyr Thr
110 115 120
Val Asn Phe Val Ser Gly Met Gln Phe Leu Ala Cys Ile Ser Thr
125 130 135
Asp Arg Tyr Trp Ala Val Thr Lys Ala Pro Ser Gln Ser Gly Val
140 145 150

Gly Lys Pro Cys Trp Val Ile Cys Phe Cys Val Trp Val Ala Ala
155 160 165
Ile Leu Leu Ser Ile Pro Gln Leu Val Phe Tyr Thr Val Asn His
170 175 180
Lys Ala Arg Cys Val Pro Ile Phe Pro Tyr His Leu Gly Thr Ser
185 190 195
Met Lys Ala Ser Ile Gln Ile Leu Glu Ile Cys Ile Gly Phe Ile
200 205 210
Ile Pro Phe Leu Ile Met Ala Val Cys Tyr Phe Ile Thr Ala Lys
215 220 225
Thr Leu Ile Lys Met Pro Asn Ile Lys Lys Ser Gln Pro Leu Lys
230 235 240
Val Leu Phe Thr Val Val Ile Val Phe Ile Val Thr Gln Leu Pro
245 250 255
Tyr Asn Ile Val Lys Phe Cys Gln Ala Ile Asp Ile Ile Tyr Ser
260 265 270
Leu Ile Thr Asp Cys Asp Met Ser Lys Arg Met Asp Val Ala Ile
275 280 285
Gln Ile Thr Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro
290 295 300
Val Leu Tyr Val Phe Met Gly Thr Ser Phe Lys Asn Tyr Ile Met
305 310 315
Lys Val Ala Lys Lys Tyr Gly Ser Trp Arg Arg Gln Arg Gln Asn
320 325 330
Val Glu Glu Ile Pro Phe Glu Ser Glu Asp Ala Thr Glu Pro Thr
335 340 345
Ser Thr Phe Ser Ile
350